Granger Causality Test in Predictive Conditional Modal Regression

Tae-Hwy Lee* Yaojue Xu†

Abstract

While a variable may not be predictable in mean using many macroeconomic and financial predictors, it may well be predictable in some quantiles especially in tails or in the mode. While the mode has its own merits relative to mean and median regressions, it has not been explored much in all disciplines. In this paper, we develop a test for Granger-Causality (GC) in the predictive regression for the conditional mode. The GC test is based on the seminal paper by Kemp and Silva (2012) and a recent paper on by Dimitriadis et al. (2019). It is in line with the recommendation of Ashley et al. (1980) to test for GC in out-of-sample prediction. We show that ENC statistic is asymptotically standard normal with zero mean under the null hypothesis of no GC in mode. Monte Carlo simulation shows ENC has a good size and power in finite samples.

Key Words: Mode, Modal midpoint, Scoring function, Elicitability, Granger Causality

1. Introduction

In recent years, the increasing number of researchers focus on the modal regression. Kemp and Silva (2012) introduce a semi-parametric conditional mode regression estimator when the dependent variable has a continuous conditional density. They also show that the proposed estimator is consistent and has a asymptotic distribution. Kemp et al. (2020) propose a semiparametric estimation of the conditional mode of a random vector that has a continuous conditional joint density. However, mode is not 1-elicitable because there doesn't exist a strictly consistent scoring function for mode. Thus, there is no possibility to compare and rank the mode forecasts of in terms of the realized score. Dimitriadis et al. (2019) formalize those ideas from Kemp and Silva (2012) and Kemp et al. (2020) in the decision theoretical framework. They define the generalized modal midpoint that is the minimizer of the expected loss function for mode and prove that the generalized modal midpoint tends to mode when the the bandwidth tends to zero. Thus, they define mode is asymptotically elicitable when the bandwidth tends to zero. In this paper, we want to test Granger-Causality (GC) for the conditional mode.

Out-of-sample forecast comparison is widely used in many fields because it is suggested to test Granger causality, which is used to determine whether some independent variables can predict the dependent variable (Ashley et al., 1980; Diebold & Mariano, 1995). Many papers focus on out-of-sample tests for equal predictive accuracy (Diebold & Mariano, 1995; Clark & McCracken, 2001; Clark & West, 2006, 2007). Diebold and Mariano (1995) introduce a Diebold-Mariano (DM) statistic for comparing predictive accuracy. DM statistic bases on loss differential between two models. However, for modal regression, the strictly consistent scoring function is Kernel density, then DM statistic is not feasible because two models have two different objective functions. Clark and McCracken (2001) and Clark and West (2006, 2007) show the ENC statistic has a zero mean under the null hypothesis for mean regression. In this paper, we develop the ENC statistic of modal regression to test out-of-sample GC in mode.

The paper is organized as follow. In section 2, we discuss the definitions and theorems of the elicitability. In section 3, we discuss the elicitability of the mode. In section 4,

*Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu
†Department of Economics, University of California, Riverside, CA 92521. E-mail: yxu122@ucr.edu

we develop the ENC statistic in the conditional modal regression. We show that the DM statistic has a bias and the ENC statistic has zero mean under the null hypothesis. In section 5, by conducting Monte Carlo simulation, we show that the ENC statistic has good size and has standard normal distribution in finite sample.

2. Elicitability

In this section, we review some definitions and theorems of the elicitability. We denote an observation domain O for $y, x \in \mathrm{O} \subseteq \mathbb{R}^{d_1+d_2}$, $d_1 = \dim(y)$ and $d_2 = \dim(x)$, the conditional distribution $F \equiv F_{Y|X}$ for Y given X. Let $\mathcal{F}$ be a class of distribution function on the observation domain O, and let A be an action domain, $\gamma \in \mathrm{A}$. We define $\Gamma : \mathcal{F} \to \mathrm{A}$ be a functional. For example, $\Gamma(F(y|x))$ may be $\mathbb{E}(Y|X)$, $Q(Y|X)$, $V(Y|X)$, $\mathrm{Mode}(Y|X)$ or $\mathrm{ES}(Y|X)$, where $\mathbb{E}(Y|X)$ is the conditional mean, $Q(Y|X)$ is the conditional quantile, $V(Y|X)$ is the conditional variance, $\mathrm{Mode}(Y|X)$ is the conditional mode and $\mathrm{ES}(Y|X)$ is the conditional Expected Shortfall. Note that Γ can be a vector of several of these.

Definition 1: (Gneiting, 2011; Fissler & Ziegel, 2016) A scoring function is an $\mathcal{F}-$integrable function $S : A \times O \to \mathbb{R}$. S is said to be *$\mathcal{F}$-consistent* for a functional $\Gamma : \mathcal{F} \to \mathrm{A}$ if $\mathbb{E}_F S(\Gamma(F), Y) \leq \mathbb{E}_F S(\gamma, Y)$ for all $F \in \mathcal{F}$ and for all $\gamma \in \mathrm{A}$. Furthermore, S is *strictly $\mathcal{F}$-consistent* for Γ if it is $\mathcal{F}$-consistent for Γ and if $\mathbb{E}_F S(\Gamma(F), Y) = \mathbb{E}_F S(\gamma, Y)$ implies that $\gamma = \Gamma(F)$ for all $F \in \mathcal{F}$ and for all $\gamma \in \mathrm{A}$. □

Definition 2: (Gneiting, 2011; Fissler & Ziegel, 2016) A functional $\Gamma : \mathcal{F} \to \mathrm{A} \subseteq \mathbb{R}^k$ is called *k-elicitable*, if there exists a strictly $\mathcal{F}$-consistent scoring function for Γ. □

A statistical functional is elicitable if there exists a scoring function that the correct forecast of the functional is unique minimizer of the expected score. We can compare or rank the forecasts of the elicitable functional with their realized scores (Fissler & Ziegel, 2016). Many statistical functionals are 1-elicitable such as expectation, ratios of expectations, quantiles (Value-at-Risk) and expectiles. However, some are not 1-elicitable such as variance, mode or Expected Shortfall (Gneiting, 2011).

3. Elicitability of Mode

In this section, we discuss scoring rule, asymptotic elicitability and identification function for the mode.

3.1 Scoring Rule for the Mode and the Modal Midpoint

The scoring function of the mode can be written as one minus kernel or minus kernel. See Lee (1989, 1993), Kemp and Silva (2012) and Dimitriadis et al. (2019).

Definition 3: (Kemp & Silva, 2012) The scoring function for the mode is

$$S_\delta^K(\gamma, y) = -\frac{1}{\delta} K\left(\frac{y-\gamma}{\delta}\right), \tag{1}$$

where δ is the strictly positive bandwidth and $K(\cdot)$ denotes a smooth kernel function. □

Definition 4: (Dimitriadis et al., 2019) $\Gamma_\delta(F)$ is the *conditional modal midpoint* that is the minimizer of the (expected) scoring function in equation (1)

$$\Gamma_\delta(F) = \arg\min_{\gamma \in \mathbb{R}} \mathbb{E}_F \left[S_\delta^K(\gamma, Y) \right]. \tag{2}$$

□

Assumption 1: F is a unimodal distribution, so that $S_\delta^K(\gamma, y)$ in equation (1) is strictly $\mathcal{F}$-consistent scoring function. □

Therefore, the conditional modal midpoint $\Gamma_\delta(F)$ is 1-elicitable. However, note that the conditional mode $\Gamma(F)$ is not elicitable.

3.2 Asymptotic Elicitability

Gneiting (2011) mentions informally that the mode is an optimal forecast under the zero-one scoring function $S_\delta(\gamma, y) = 1(|\gamma - y| > \delta)$. $\Gamma_\delta(F)$ is defined as the modal midpoint functional. The scoring function $S_\delta(\gamma, y)$ is consistent for $\Gamma_\delta(F)$. Thus, the mode functional can be defined as $\Gamma(F) = \lim_{\delta \to 0} \Gamma_\delta(F)$ (Gneiting, 2011). Dimitriadis et al. (2019) formalize the idea from Gneiting (2011), Kemp and Silva (2012) and Kemp et al. (2020) and introduce the concept of a "asymptotic elicitability".

Definition 5: (Gneiting, 2011; Dimitriadis et al., 2019) The functional $\Gamma : \mathcal{F} \to A \subseteq \mathbb{R}$ is *asymptotically elicitable* if there exists a sequence of elicitable functional $\Gamma_\delta : \mathcal{F} \to A \subseteq \mathbb{R}$ such that $\lim_{\delta \to 0} \Gamma_\delta(F) = \Gamma(F)$ for all $F \in \mathcal{F}$. □

The following theorem shows that the mode is asymptotically elicitable when the bandwidth δ tends to 0.

Theorem 1: (Dimitriadis et al., 2019) For the class of distributions $\mathcal{F}$ consists of absolutely continuous unimodal distributions with bounded density and for any kernel function K which is positive, smooth, $\int K(u)\mathrm{d}u = 1$ and $\log(K(u))$ is a concave function, the functional Γ_δ induced by the scoring function (1) is well-defined for all $\delta > 0$, it holds that

$$\lim_{\delta \to 0} \Gamma_\delta(F) = \Gamma(F), \tag{3}$$

for all $F \in \mathcal{F}$, where $\Gamma(F)$ is the conditional mode of the conditional distribution $F \equiv F_{Y|X}$. □

The mode is not 1-elicitable but 1-elicitable when the bandwidth δ tends to 0, i.e., the mode is asymptotically 1-elicitable.

4. Granger-Causality Test in Predictive Modal Regression

In this section, we discuss the estimation of the modal regression and develop a new test for Granger Causality in the conditional modal regression.

4.1 Models

We have two models. One model is without conditioning on x and the other model is with conditioning on x. The unconditional mode of the unconditional distribution $F_1 = F_Y(Y)$

is $\Gamma(F_1)$. The conditional mode of the conditional distribution $F_2 = F_{Y|X}(Y|X)$ is the functional $\Gamma(F_2)$. The two nested mode models are

$$\text{Model 1} : y_{t+1} = a_1 + u_{t+1}^{(1)} \equiv x_{1,t}'\beta_1 + u_{t+1}^{(1)}, \tag{4}$$

$$\text{Model 2} : y_{t+1} = a_{2,} + bx_t + u_{t+1}^{(2)} \equiv x_{2,t}'\beta_2 + u_{t+1}^{(2)}, \tag{5}$$

where $\gamma_t^{(1)} = x_{1,t}'\beta_1$ and $\gamma_t^{(2)} = x_{2,t}'\beta_2$.The dependent variable y_{t+1} is a scalar random variable. The independent variable x_t is stationary variable. $x_{1,t}$ is a strict subset of $x_{2,t}$. $x_{1,t}' = 1, \beta_1 = a_1, x_{2,t}' = (1, x_t), \beta_2 = (a_2, b)$. Thus, we estimate $\hat{u}_{t+1}^{(1)}$ and $\hat{u}_{t+1}^{(2)}$ by following $\hat{u}_{t+1}^{(1)} = y_{t+1} - \hat{a}_{1,t}$ and $\hat{u}_{t+1}^{(2)} = y_{t+1} - \hat{a}_{2,t} - \hat{b}_t x_t$.

4.2 Estimation of Modal Regression

The scoring function of mode includes the Kernel function, so we need to choose the bandwidth δ_R. Following Kemp and Silva (2012), we define

$$\delta_n = k \, \text{MAD} \, n^{-1/7}. \tag{6}$$

For the value of k, we choose $k = 0.8, 1.6$. According to Kemp and Silva (2012), $k = 1.6$ is inspired by Silverman (1986) rule-of-thumb. $k = 0.8$ is chosen to show under-smoothing. MAD denotes the median of the absolute deviation from the median ordinary least squares residual,

$$\text{MAD} = \text{med}_t \left[\text{abs} \left((y_{t+1} - x_{i,t}'\beta_i) - \text{med}_t \left(y_{t+1} - x_{i,t}'\beta_i \right) \right) \right], \tag{7}$$

where β is the ordinary least squares (OLS) estimator.

In this paper, we choose the standard normal density kernel, that is

$$K\left(\frac{y-\gamma}{\delta}\right) = \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{(y-\gamma)^2}{2\delta^2}\right). \tag{8}$$

In order to find out the moment condition, taking the derivative of the expectation of equation (1) with respect to β_i, we get

$$\frac{\partial \mathbb{E}_F S_\delta^K(\gamma, Y)}{\partial \beta_i} = \mathbb{E}_F \left[\sum_{t=1}^{n} \exp\left(-\frac{\left(y_{t+1} - x_{i,t}'\beta_i\right)^2}{2\delta_n^2} \right) \left(y_{t+1} - x_{i,t}'\beta_i \right) x_{i,t}' \right] = 0.$$

To find out the estimators, we solve the following moment condition

$$\mathbb{E}_F \left[\sum_{t=1}^{n} w_t(\beta_i) \left(y_{t+1} - x_{i,t}'\beta_i \right) x_{i,t}' \right] = 0, \tag{9}$$

where the weight is

$$w_t(\beta_i) = \exp\left(-\frac{\left(y_{t+1} - x_{i,t}'\beta_i\right)^2}{2\delta_n^2} \right). \tag{10}$$

Solving the moment equation (9) gives the estimator of the mode regression when δ_n tends to zero. Solving the moment equation (9) gives the estimator of the mean regression when

δ_n tends to ∞. The estimator $\hat{\beta}_i$ can be computed as iterated weighted least squares estimators, i.e., as the solution to the equation:

$$\hat{\beta}_i = \left[\sum_{t=1}^{n} w_t \left(\hat{\beta}_i \right) x_{i,t} x'_{i,t} \right]^{-1} \sum_{t=1}^{n} w_t \left(\hat{\beta}_i \right) x_{i,t} y_{t+1}, \tag{11}$$

where $w_t\left(\hat{\beta}_i\right) = \exp\left(-\frac{\left(y_{t+1} - x'_{i,t}\hat{\beta}_i\right)^2}{2\delta_n^2} \right)$.

Taking the derivative of scoring function with respect to $\beta_{i,t}$, we get the score [1]

$$h_{i,t} = \frac{\partial S_\delta^K(\gamma, y)}{\partial \beta_{i,t}} = -\frac{1}{\delta_n^2} K' \left(\frac{u_t^{(i)}}{\delta_n} \right) x_{i,t-1}. \tag{12}$$

Then, let $H_i(t)$ as follows:

$$H_i(t) = \frac{1}{R} \sum_{j=t-R+1}^{t} h_{i,t} = -\frac{1}{R} \sum_{j=t-R+1}^{t} \frac{1}{\delta_R^2} K' \left(\frac{u_t^{(i)}}{\delta} \right) x_{i,t-1}. \tag{13}$$

Taking the derivative of $h_{i,t}$ with respect to $\beta_{i,t}$, the Hessian is obtained as follows:

$$\Lambda_{i,t} = \frac{\partial h_{i,t}}{\partial \beta} = \frac{1}{\delta_n^3} K'' \left(\frac{u_t^{(i)}}{\delta_n} \right) x_{i,t-1} x'_{i,t-1}. \tag{14}$$

According to Kemp and Silva (2012), we get that

$$\frac{1}{n} \sum_{t=1}^{n} \Lambda_{i,t} = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\delta_n^3} K'' \left(\frac{\hat{u}_t^{(i)}}{\delta_n} \right) x_{i,t-1} x'_{i,t-1} = B_i^{-1} + o_p(1), \tag{15}$$

where

$$B_i^{-1} = \mathbb{E}_F \left[f''_{u_t^{(i)}|X} \left(0 | x_{i,t} \right) x_{i,t} x'_{i,t} \right] = \lim_{n \to \infty} \mathbb{E}_F \left[\frac{1}{\delta_n^3} \left\{ K'' \left(\frac{u_t^{(i)}}{\delta_n} \right) \right\} x_{i,t} x'_{i,t} \right], \tag{16}$$

where $f''_{u^{(i)}|X}\left(0|x_t\right) = \left.\frac{\partial^2 L}{\partial\beta\partial\beta}\right|_{\beta_0}$ and $\mathbb{E}_F f''_{u^{(i)}|X}\left(0|x_t\right)$ is negative definite. Kemp and Silva (2012) shows that the asymptotic distribution of $\hat{\beta}_{i,t}$ is

$$\sqrt{n\delta_n^3} \left(\hat{\beta}_{i,t} - \beta_i \right) = -B_i^{-1} \left[\frac{1}{\sqrt{n\delta_n}} \sum_{t=1}^{n} K' \left(\frac{u_{i,t}}{\delta_n} \right) x_{i,t} \right] + o_p(1) \xrightarrow{d} N\left(0, B_i^{-1} A_i B_i^{-1}\right), \tag{17}$$

where $A_i = \lim_{n\to\infty} \mathbb{E}_F \left[\frac{1}{\delta_n} \left\{ K'\left(\frac{u_i}{\delta_n}\right) \right\}^2 x_i x'_i \right]$. Equation (17) shows that the mode regression estimator converges to a normal distribution.

4.3 A New Test for Granger Causality in Modal Regression

In this paper, we develop a new test for Grander Causality based on the combined mode

$$\gamma^{(c)} = (1 - \lambda)\gamma^{(1)} + \lambda\gamma^{(2)}. \tag{18}$$

[1] We note that $h_{i,t}$ is called the score (the first order condition of $S(\gamma, Y)$), while $S(\gamma, Y)$ is called the scoring function or scoring rule.

The null and alternative hypotheses are

$$\mathbb{H}_0 : \lambda = 0, \text{ and } \mathbb{H}_1 : \lambda \neq 0. \tag{19}$$

Under $\mathbb{H}_0 : \lambda = 0$, there is no Granger Causality between X and Y. Under $\mathbb{H}_1 : \lambda \neq 0$, there exists Granger Causality between X and Y.

Denote the ENC statistic as $\hat{C}_P$. Under $\mathbb{H}_0$,

$$C \equiv \mathbb{E}_F \left[\frac{1}{\delta_R^2} K' \left(\frac{u_{t+1}^{(1)}}{\delta_R} \right) \left(u_{t+1}^{(1)} - u_{t+1}^{(2)} \right) \right] = 0, \tag{20}$$

$$\hat{C}_P \equiv P^{-1} \sum_{t=R}^{T} \left[\frac{1}{\delta_R^2} K' \left(\frac{\hat{u}_{t+1}^{(1)}}{\delta_R} \right) \left(\hat{u}_{t+1}^{(1)} - \hat{u}_{t+1}^{(2)} \right) \right] \xrightarrow{p} C = 0 \tag{21}$$

as $P \to \infty$. □

We compare the DM statistic and the ENC statistic with the CCS statistic that is the test statistic by Chao et al. (2001). Chao et al. (2001) show that the CCS statistic is $\hat{M}_P = P^{-1} \sum_{t=R}^{T} \hat{u}_{t+1}^{(1)} x_t$. In order to compare the equal predictive accuracy of two nested mode models, we standardize these three statistics. The DM statistic is $DM_P \equiv \hat{S}_P^{-0.5} \sqrt{P} \hat{D}_P$, where $S_P = \text{var}\left(\sqrt{P}\hat{D}_P\right)$ and $S_P - \hat{S}_P \xrightarrow{p} 0$. The ENC statistic is $ENC_P \equiv \hat{Q}_P^{-0.5} \sqrt{P} \hat{C}_P$, where $Q_P = \text{var}\left(\sqrt{P}\hat{C}_P\right)$ and $Q_P - \hat{Q}_P \xrightarrow{p} 0$. The CCS statistic is $CCS_P \equiv \hat{W}_P^{-0.5} \sqrt{P} \hat{M}_P$, where $W_P = \text{var}\left(\sqrt{P}\hat{M}_P\right)$ and $W_P - \hat{W}_P \xrightarrow{p} 0$.

It can be shown that ENC_P is asymptotically standard normal under the null hypothesis. Its proof is to be presented in the full version of the paper. Below, we examine the finite sample properties of the ENC statistic, in comparison with the DM and CCS statistics. The Monte Carlo results in the next section confirms that ENC_P is asymptotically standard normal under the null hypothesis, has a proper size and excellent power, while DM has a size problem and CCS has a power problem.

5. Monte Carlo Simulation

In this section, we show that the ENC_P has good size and the distribution of ENC_P is standard normal by using Monte Carlo Simulation.

5.1 Simulation Design

In order to show the distribution of ENC statistic, we simulate data from the following DGP. We generate the additional variable x_t in Model 2 to be an AR(1) process. We set

$$x_t = \phi x_{t-1} + v_t, \tag{22}$$

where $|\phi| < 1$, $v_t \sim N(0, \sigma_v^2)$. Following Dimitriadis et al. (2019), we generate the error term $u_{t+1} \overset{iid}{\sim} SN(0, \sigma_u^2, \eta)$, where $SN(0, \sigma_u^2, \eta)$ is a skewed normal distribution and η is the skewness of the error term u.

Following Kemp and Silva (2012), we define the bandwidth δ_R as

$$\delta_R = k \, \text{MAD} \, R^{-1/7}. \tag{23}$$

For the value of k, we choose $k = 0.8, 1.6$. According to Kemp and Silva (2012), $k = 1.6$ is inspired by Silverman (1986) rule-of-thumb. $k = 0.8$ is chosen to show under-smoothing.

Table 1: Size of Test: $k = 1.6$, $b = 0$, $\phi = 0$, $\sigma_u = 1$, at 5% nominal level, 2000 Repeats

Repeat=2000		$P = 240$ DM$_P$	$P = 240$ ENC$_P$	$P = 240$ CCS$_P$	$P = 480$ DM$_P$	$P = 480$ ENC$_P$	$P = 480$ CCS$_P$
$\eta = 0.25$	$R = 240$	0.005	0.046	0.050	0.003	0.049	0.049
	$R = 480$	0.008	0.044	0.053	0.004	0.041	0.047
$\eta = 0.5$	$R = 240$	0.010	0.042	0.055	0.004	0.049	0.054
	$R = 480$	0.014	0.043	0.056	0.005	0.040	0.048

Table 2: Power of Test: $k = 1.6$, $b = 0.1$, $\phi = 0$, at 5% nominal level, 2000 Repeats

Repeat=2000		$P = 240$ DM$_P$	$P = 240$ ENC$_P$	$P = 240$ CCS$_P$	$P = 480$ DM$_P$	$P = 480$ ENC$_P$	$P = 480$ CCS$_P$
$\eta = 0.25$	$R = 240$	0.031	0.125	0.051	0.037	0.197	0.053
	$R = 480$	0.039	0.122	0.056	0.043	0.235	0.052
$\eta = 0.5$	$R = 240$	0.029	0.115	0.052	0.022	0.150	0.050
	$R = 480$	0.046	0.103	0.056	0.030	0.134	0.053

MAD denotes the median of the absolute deviation from the median ordinary least squares residual,

$$\text{MAD} = \text{med}_t \left[\text{abs} \left(\left(y_{t+1} - x'_{i,t}\beta_i \right) - \text{med}_t \left(y_{t+1} - x'_{i,t}\beta_i \right) \right) \right], \tag{24}$$

where β is the OLS estimator.

We consider $\alpha \in \{5, 0.05\}$, $\eta \in \{0.25, 0.5\}$, $\phi \in \{0\}$, $\sigma_v \in \{1\}$, $\sigma_u \in \{1\}$, $c_2 \in \{0.5\}$, and $b \in \{0, 0.1\}$.

We use the MATLAB to estimate and forecast the models. We choose to use standard normal density Kernel. To find out the estimators, Kemp and Silva (2012) point out that we can solve the following moment condition

$$\mathbb{E}_F \left[\sum_{t=1}^{R} w_t \left(\beta_i \right) \left(y_{t+1} - x'_{i,t}\beta_i \right) x'_{i,t} \right] = 0, \tag{25}$$

where the weight $w_t \left(\beta_i \right) = \exp \left(-\frac{\left(y_{t+1} - x'_{i,t}\beta_i \right)^2}{2\delta_R^2} \right)$. The equation (25) is for the mode regression when δ_R tends to zero. The equation (25) is for the mean regression when δ_R tends to ∞. The estimator $\hat{\beta}_i$ can be computed as iterated weighted least squares estimators, i.e, as the solution to the equation:

$$\hat{\beta}_i = \left[\sum_{t=1}^{R} w_t \left(\hat{\beta}_i \right) x_{i,t} x'_{i,t} \right]^{-1} \sum_{t=1}^{R} w_t \left(\hat{\beta}_i \right) x_{i,t} y_{t+1}, \tag{26}$$

where $w_t \left(\hat{\beta}_i \right) = \exp \left(-\frac{\left(y_{t+1} - x'_{i,t}\hat{\beta}_i \right)^2}{2\delta_R^2} \right)$.

For Model 1, we regress $\{y_s\}_{s=t-R+1}^{t}$ on constant term to get $\hat{c}_{1,t}$, where $t = R, \cdots, T$. For Model 2, we regress $\{y_s\}_{s=t-R+1}^{t}$ on $\{1, x_{s-1}\}_{s=t-R+1}^{t}$ to get $\hat{a}_{2,t}$ and $\hat{b}_t$. The forecast errors are $\hat{u}_{t+1}^{(1)} = y_{t+1} - \hat{a}_{1,t}$ for Model 1 and $\hat{u}_{t+1}^{(1)} = y_{t+1} - \hat{a}_{2,t} - \hat{b}_t x_t$, where

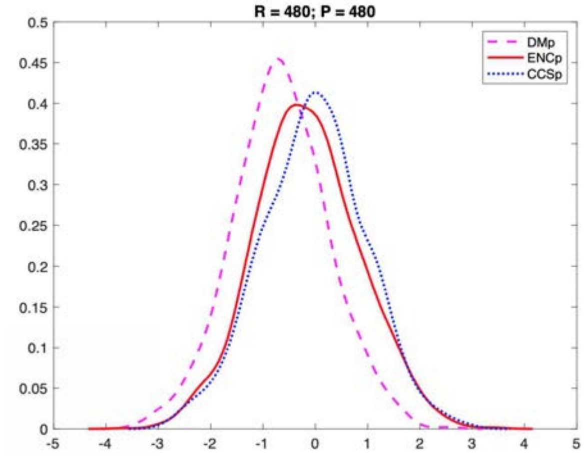

Figure 1: The distribution of DM_P, ENC_P, and CCS_P, $k = 1.6$, $\eta = 0.5$, $b = 0$, $\phi = 0$, 2000 Repeats..

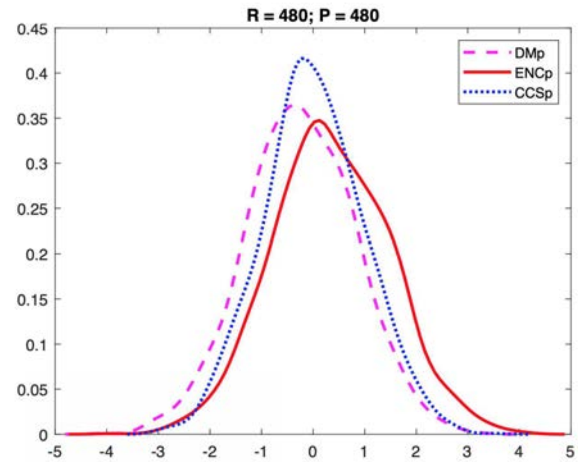

Figure 2: The distribution of DM_P, ENC_P, and CCS_P, $k = 1.6$, $\eta = 0.5$, $b = 0.1$, $\phi = 0$, 2000 Repeats..

$t = R, \cdots, T$. We choose to use rolling window for estimation. The in-sample observations $R \in \{240, 480\}$ and the out-of-sample observations $P \in \{240, 480\}$. We repeat 2000 times to find out the distributions of DM_P, ENC_P and CCS_P.

5.2 Simulation Results

Figures 1-2 and Tables 1-2 show the Monte Carlo simulation distribution and the size and power of the DM_P, ENC_P and CCS_P statistics with different skewness. Table 1 shows the size of the test under $\mathbb{H}_0$ with different skewness η. This table demonstrates that DM_P is much less than 5% under 5% nominal level, which means that the DM_P has a bias under the null hypothesis. However, comparing to the DM_P, the size of the ENC_P and the CCS_P is good under the 5% nominal level. Table 2 shows the the power of test under $\mathbb{H}_1$ with different skewness η. This table demonstrates that the ENC_P has the highest power for different in-sample observations R, out-of-sample forecasts P and skewness η. Moreover, the power of the CCS_P is lower than that of the ENC_P but is higher than the power of the DM_P.

Figure 1 shows the distributions of the DM_P, ENC_P and CCS_P statistics with the skewness $\eta = 0.5$ under $\mathbb{H}_0$. Figure 1 demonstrates that the $\mathrm{DM_P}$ has the negative mean and high kurtosis, which also implies that the DM_P has a downward bias under $\mathbb{H}_0$. The distributions of the ENC_P, and the $\mathrm{CCS_P}$ are close to the standard normal distribution under the null hypothesis. Figure 2 shows the asymptotic distribution of the DM_P, $\mathrm{ENC_P}$ and CCS_P statistics with the skewness $\eta = 0.5$ under $\mathbb{H}_1$. From Figure 2, we can see that the mean of the DM_P is lower than means of the ENC_P and the CCS_P.

6. Conclusion

In this paper, we develop a new statistic for Granger-Causality test in the conditional modal regression. The scoring function for the mode is based on the papers by Kemp and Silva (2012) and Dimitriadis et al. (2019). We show that the ENC statistic has zero mean under the null hypothesis of no GC in the mode. Monte Carlo simulation demonstrates that the ENC statistic has good size and power and has standard normal distribution in finite samples.

References

Ashley, R., Granger, C. W., & Schmalensee, R. (1980). Advertising and aggregate consumption: An analysis of causality. *Econometrica*, *48*(5), 1149–1167.

Chao, J., Corradi, V., & Swanson, N. R. (2001). Out-of-sample tests for granger causality. *Macroeconomic Dynamics*, *5*(4), 598–620.

Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, *105*(1), 85–110.

Clark, T. E., & West, K. D. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, *135*(1-2), 155–186.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291–311.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.

Dimitriadis, T., Patton, A. J., & Schmidt, P. (2019). Testing forecast rationality for measures of central tendency. *arXiv:1910.12545*.

Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, *44*(4), 1680–1707.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.

Kemp, G. C., Parente, P. M., & Silva, J. S. (2020). Dynamic vector mode regression. *Journal of Business & Economic Statistics*, *38*(3), 647–661.

Kemp, G. C., & Silva, J. S. (2012). Regression towards the mode. *Journal of Econometrics*, *170*(1), 92–101.

Lee, M.-J. (1989). Mode regression. *Journal of Econometrics*, *42*(3), 337–349.

Lee, M.-J. (1993). Quadratic mode regression. *Journal of Econometrics*, *57*(1-3), 1–19.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis* (Vol. 26). CRC press.